JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1977, 10, 325-332 NUMBER 2 (SUMMER) 1977

TAKING A CLOSER LOOK: TIME SAMPLING AND MEASUREMENT ERROR[1]

J. POWELL, B. MARTINDALE, S. KULP, A. MARTINDALE, AND R. BAUMAN

CALIFORNIA STATE COLLEGE, CALIFORNIA, PENNSYLVANIA

A person manufactured his in-seat behavior for 15, 30-min sessions so that there were three blocks of five sessions where the behavior occurred 20%, 50%, and 80% of the time. Whole interval, partial interval, and momentary time-sample measures of the behavior were taken and compared to the continuous measure of the behavior *i.e.,* per cent of time the behavior occurred. For interval time sampling, the difference between the continuous and sample measures *i.e.,* measurement error, was: (1) extensive, (2) unidirectional, (3) a function of the time per response, and (4) inconsistent across changes in the continuous measure. A procedural analysis demonstrated that the frequency and duration of behavior are confounded in interval time sampling. Momentary time sampling was found to be superior to interval time sampling in estimating the duration a behavior occurs.

DESCRIPTORS: time sampling, partial interval, whole interval, momentary, measurement error

Powell, Martindale and Kulp, (1975) assessed the validity of time-sample measures of behavior by comparing the percentage of time that a behavior occurred *i.e.,* a continuous measure, with the percentage of observed occurrences of the behavior *i.e.,* a sample measure. The findings were: (1) requiring any instance of behavior within an interval to score that interval (partial-interval time sampling) overestimated the continuous measure, (2) requiring the behavior to last throughout the interval to score that interval (whole interval time sampling) both over- and underestimated the continuous measure, and (3) requiring that the behavior be observed exactly at the end of the interval to score that interval (momentary time sampling) both over- and underestimated the continuous measure. The difference between the continuous and sample measures, measurement error, was a function of the frequency of the sample measures.

The present investigation explored, more completely, the direction and extent of measurement error in time sampling. Partial interval, whole interval, and momentary time samples were conducted in all sessions so that comparisons could be made among the three procedures. The total time the behavior occurred was varied to determine how changes in the continuous measure influenced the sample measures. Also, the data were analyzed to determine whether, as previously predicted, that measurement error in interval time sampling would be (1) a function of the time per response of the observed behavior and (2) inconsistent across experimental conditions. The intent of the research was to provide objective information that investigators can consider in selecting appropriate measurement parameters.

METHOD

An adult male manufactured his in-seat behavior *i.e.,* posterior in contact with the seat of the chair, for 15, 30-min sessions. For five sessions each, this person was instructed to be in his seat 6 min, 15 min, and 24 min. This

[1]A much abbreviated and earlier version of this paper was presented at the American Psychological Association Convention, Chicago, 1975. Reprints may be obtained from J. Powell, Department of Special Education, California State College, California, Pennsylvania 15419.

gave three blocks of five sessions where the behavior occurred 20%, 50%, and 80% of the time. The person controlled his in-seat behavior by observing two running-time meters. One meter can continuously and served to determine the 30-min session length; the second ran when the person operated a switch, which he did whenever he was seated. All sessions were videotaped and a 50-msec tone that sounded each 5 sec was imposed on the recordings. The tapes were then reviewed and the status of the behavior assessed within and at the end of each 5-sec interval. Within each interval, the behavior was recorded as occurring the entire interval, a part of the interval, or none of the interval. At the end of each interval, the behavior was recorded as occurring or not occurring. Next, the data sheets were examined to determine the effect of varying the length of the observation interval. For example, 10-sec interval recording simply required considering the 5-sec intervals in pairs, and 10-sec momentary time sampling required noting the state of the behavior at the end of every other interval. Results were obtained for interval time sampling when the interval length was 5, 10, 20, 60, 120, 200, and 300 sec, *i.e.*, the number of observations per session ranged from 360 to six. Results were obtained for momentary time sampling when the interval length was 5, 10, 20, 60, 120, 200, 300, 600, 900, and 1800 sec, *i.e.*, the number of observations ranged from 360 to one.

All videotapes were independently reviewed to determine an index of interobserver agreement. For interval time sampling, this measure was calculated for entire, partial, and no intervals of behavior; for momentary time sampling, this measure was calculated for occurrences and nonoccurrences of the behavior. The mean percentage agreement for the interval measures was 96%, 99%, and 99% and for the momentary measures, 99% and 99%. Also recorded were the number of times the person sat down, and, using a running-time meter, the total time the behavior occurred. There were no disagreements on the number of responses, and agreement on the total time always exceeded 99%.

RESULTS

Figure 1 shows, as a function of the length of the observation interval, the per cent of observations scored for whole-interval, partial-interval,[2] and momentary time sampling. This figure shows that whole and partial interval time sampling, respectively, under- and overestimated the continuous measure: the magnitude of this error was a function of the observation length. This figure demonstrates how differences in the time the behavior occurred influenced the interval sampling error. When the behavior occurred 20% of the time, the maximum possible difference between the continuous and partial interval measure was 80%; this difference was reached when the observation interval was 200 sec. When the behavior occurred 80% of the time, this maximum possible difference was only 20%; this difference was reached when the observation interval was 120 sec. Momentary time sampling both under- and overestimated the continuous measure, with a minimum difference of 20% between the two measures first observed when the time between observations reached 200 sec. At those observation lengths beyond 5 sec where both momentary and interval measures were made *i.e.*, 10, 20, 60, 120, 200, and 300 sec, the momentary measure was superior in estimating the per cent of time the behavior occurred. A comparison of the momentary and interval measures at these observation lengths shows that the momentary measure more closely approximated the continuous measure on 53 of 54 occasions. When the observation length was 5 sec, the three measurement procedures were

[2]In partial-interval time sampling, any instance of behavior within an interval results in scoring that interval. This condition is satisfied by both those intervals where the behavior occurs part of the interval and those intervals where the behavior occurs throughout the interval.

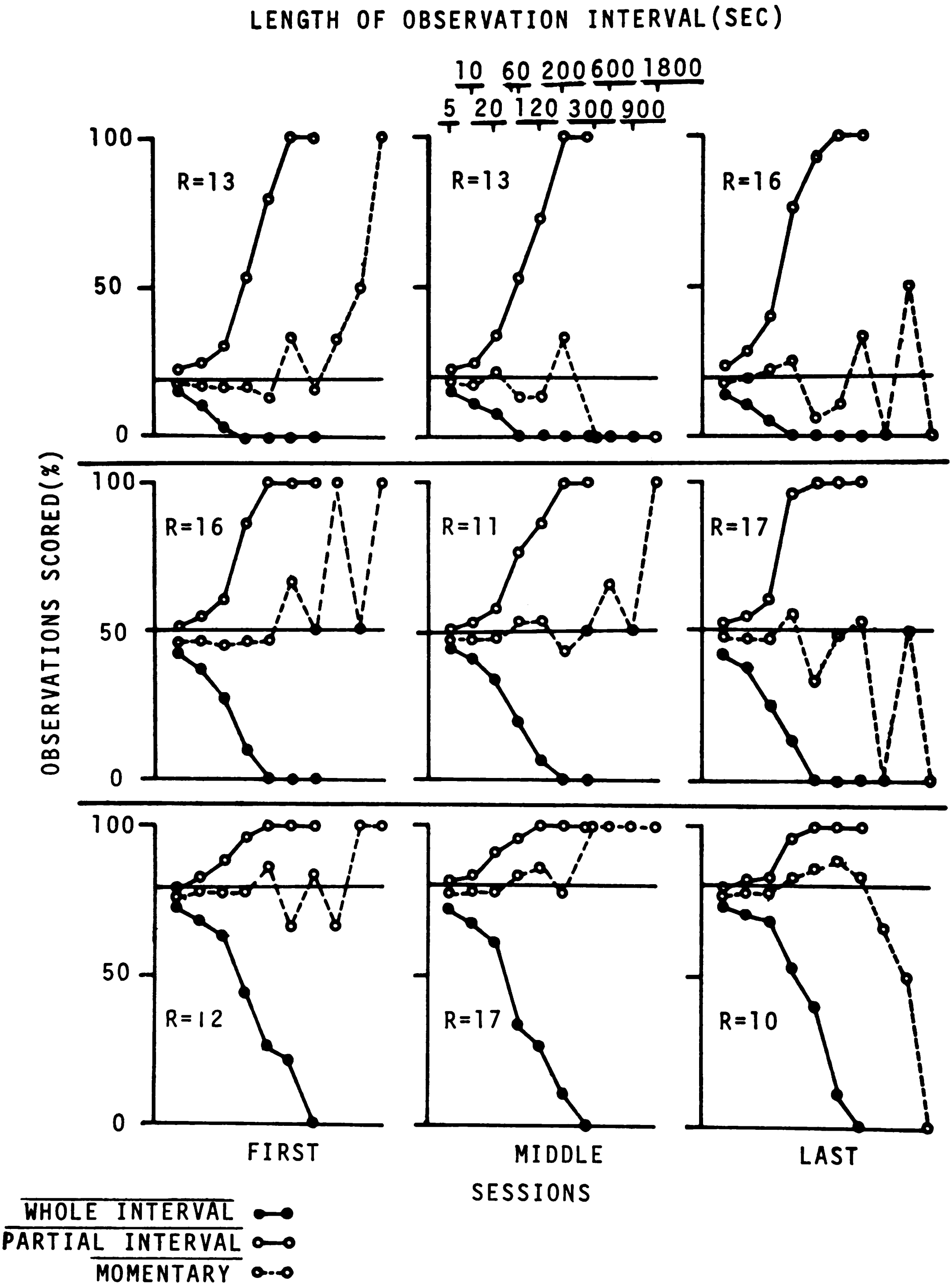


Fig. 1. Per cent of observations scored as a function of the length of the observation interval. The first, middle, and last session are shown when in-seat behavior occurred 20% (top panel), 50% (middle panel), and 80% (bottom panel) of the 30-min sessions. The number of times the person sat down (R) is given for each session.

equally effective in describing the continuous measure.

Figure 2 shows the same information as the previous figure but includes the data from all 15 sessions. The upper abscissa in this figure shows the number of observations per session for a given observation length. Figure 2 shows that the error in both interval sampling procedures increased as the number of observations per session decreased. The "smoothness" of these functional relationships is such that interpolation along any value of the independent variable results in an accurate prediction of the percentage of intervals scored. To substantiate the predictive power of these functions, Figure 2 was visually examined, and the per cent of intervals scored for 30-sec partial interval time sampling was estimated when the behavior occurred 20%, 50%, and 80% of the time. These three values were 38%, 67%, and 90%. The original data sheets were then reviewed and the actual percentage of intervals scored across the three conditions was calculated. These percentages were 37.7%, 69%, and 92.3%. The predicted values did not deviate from the actual values by more than 2.5%. Again, Figure 2 shows how changes in the continuous measure influenced the interval sampling error. When the behavior occurred 80% of the time, the error in partial interval time sampling was restricted, but the error in whole interval time sampling was essentially unrestricted. Conversely, when the behavior occurred 20% of the time the error in the partial interval measure was large, but the error in the whole interval measure was small. Figure 2 was examined to determine the observation length where a minimum difference of 20% between the interval and continuous measure was first observed. For partial interval time sampling, these durations were 60, 60, and 120 sec when the behavior occurred 20%, 50%, and 80% of the time. For whole interval time sampling, they were 60, 60, and 60 sec. Figure 2 shows that the effect of grouping the momentary time-sampling data was to produce functions that closely approached the continuous measure. Only three of the 30 data points deviated 20% or more from the continuous measure. Two of these points were when the time between observations was 1800 sec, that is, when one observation per session occurred; the third point was when the time between observations was 900 sec, that is, when two observations per session occurred.

The data in Figure 2 were used to determine how interval sampling is influenced by manipulated changes in the duration of behavior. This was accomplished by comparing differences in the continuous measure with the corresponding differences in the interval measure (see Figure 3). Figure 3 demonstrates that the magnitude of the behavior change, as expressed by the continuous measure, exceeded the change expressed by the interval measure. Also, as the observation interval increased, the ability of the interval measure to reflect changes in the continuous measure decreased. In fact, when the observation interval reached 120 sec (top panel), and 200 sec (bottom panel), the partial interval measure revealed no change.

Scattergrams (not shown) were constructed relating the time per response of the behavior to the measurement error in the partial-interval measure. For all sessions, the time per response was calculated by dividing the total time the behavior occurred by the number of in-seat responses. The measurement error equalled the difference between the per cent of intervals scored and the per cent of time the behavior occurred. This error was determined for observation lengths of 5, 10, 20, 60, 120, and 300 sec. The ellipitical patterning of the scattergrams justified a mathematical description of this relationship, and these results are contained in Figure 4. The lines in this figure are lines of best fit calculated by the method of least squares. This figure shows that as the time per response decreased the measurement error increased. The slopes of the straight-line equations indicate that the rate of change in the error-time per response relationship was a function of the length of the observation interval.

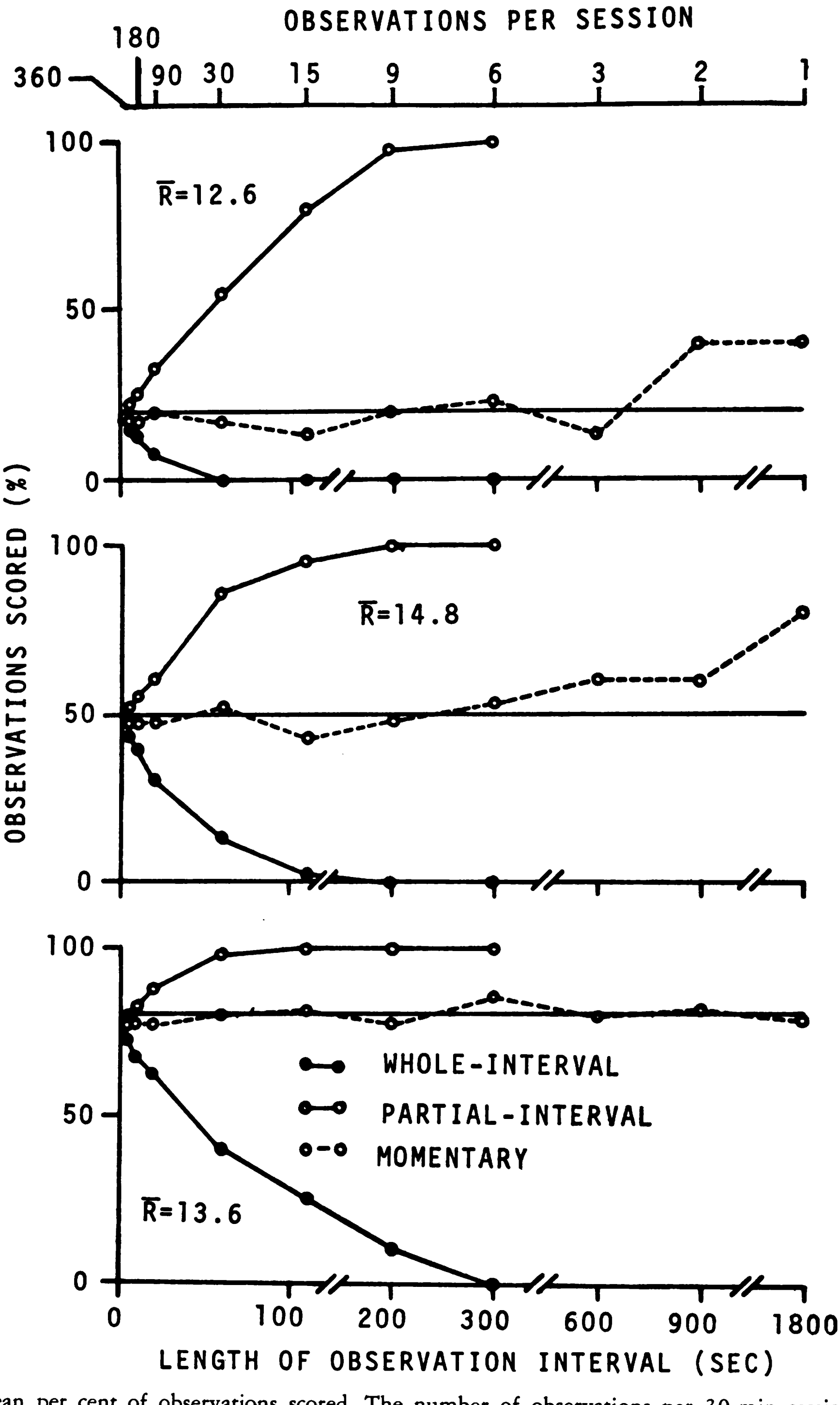


Fig. 2. Mean per cent of observations scored. The number of observations per 30-min session is shown (upper abscissa) for each observation interval (lower abscissa). These functions are shown when in-seat behavior occurred 20% (top panel), 50% (middle panel), and 80% (bottom panel) of the time. Each point represents the mean value for five sessions. R is the mean number of times the person sat down in five sessions.

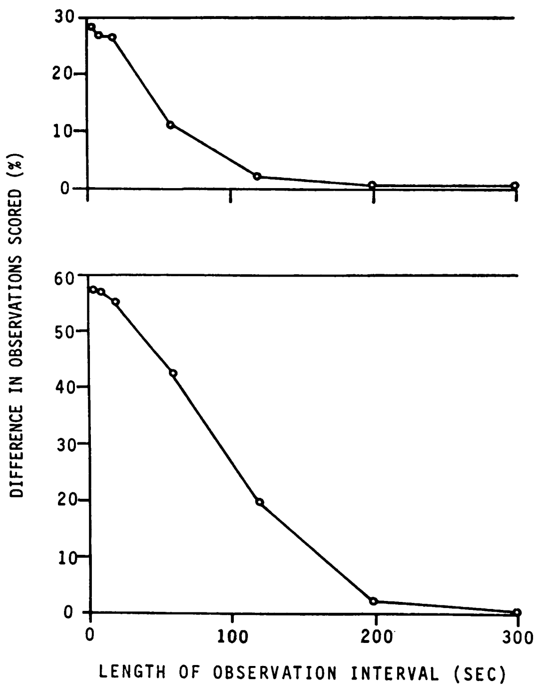


Fig. 3. The effect on partial interval time sampling of changes in the duration a behavior occurs. In the top panel, the difference of 30% (horizontal line) represents a change in the continuous measure from 80% to 50%. In the bottom panel, the difference of 60% (horizontal line) represents a change in the continuous measure from 80% to 20%. The corresponding difference in the per cent of observations scored for the partial-interval measures is shown as a function of the length of the observation interval. The data in this figure are drawn directly from Figure 2. Each point represents the mean value for five sessions.

DISCUSSION

For interval sampling, the per cent of observations scored overestimated or underestimated the amount of time the behavior occurred. This unidirectionality of the error in interval sampling means that sampling over a long period of time so as to reduce irregularities (Hutt and Hutt, 1970, p. 71) is not applicable when this measurement procedure is used.

The error in interval time sampling was large even when the observations were conducted frequently *e.g.,* the relative percentage error for 30-sec partial interval time sampling was 88%, 38%, and 15% when the behavior occurred 20%, 50%, and 80% of the time. Many inves-

$$\frac{\%\text{ of intervals} - \%\text{ continuous} \times 100}{\%\text{ continuous}}$$

tigations have employed observation intervals well beyond 30 sec (*e.g.,* Barrish, Saunders, and Wolf, 1969; Hauserman, Walen and Behling, 1973; Milby, 1970).

The above relative error figures demonstrate another finding. Namely, that interval sampling does not yield a consistent error when behavior is manipulated. Consider the interval sampling of a behavior that, via an intervention, changes from occurring often to seldom. If partial interval (whole interval) time sampling is employed, the error in the first instance will be small (large), and in the second instance large (small). Although interval error will not be consistent across experimental conditions, this study showed that changes in the continuous measure exceeded the changes expressed by the interval measure. This suggests that the magnitude of the behavioral change in many studies that have employed interval sampling was actually greater than reported.

Interval measurement has been reported as reflecting the response dimensions of frequency (Arrington, 1943), duration (Hutt and Hutt,

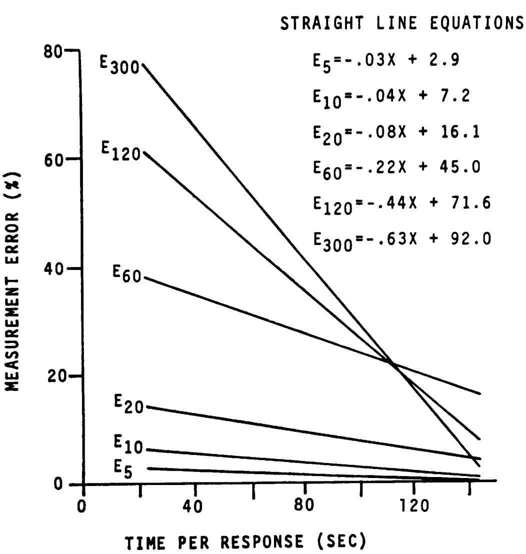
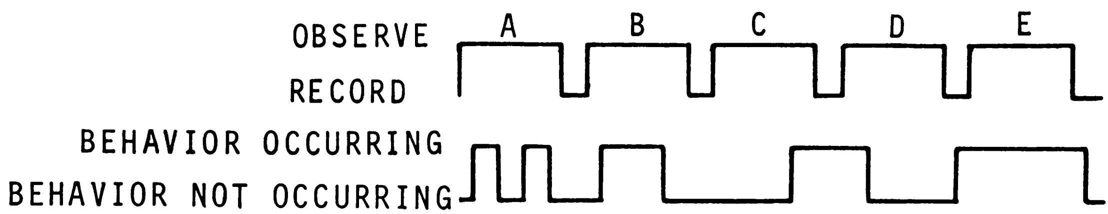


Fig. 4. The straight lines and their equations relating the time per response of the behavior to the measurement error in partial interval time sampling.

1970), or both frequency and duration (Hall, 1971). An analysis of interval sampling demonstrates that the results are not directly related to either the frequency or duration of the observed behavior.

Consider five observation intervals, "A" through "E", and an event recording of a behavior that occurs 50% of the total observation period.

behavior: a variable composed of both duration and frequency. The smallest time per response observed in this study (20 sec) is still large, compared to the expected values of many behaviors that have been subjected to interval measurement *e.g.,* mother-infant glances, talkouts. When the time per response does become small, a large increment in the error in interval sampling can be expected. For example, the time



In "A", there are multiple examples of the behavior; in "B", a single example of the behavior; in "C", the behavior is initiated, but not terminated; in "D", the behavior is not initiated, but is terminated, and in "E", the behavior occurs throughout the interval. With the exception of an interval where no behavior occurs, combinations of the above represent the total possible arrangement of behavior within intervals. Partial and whole interval time sampling in the above case would report that the behavior occurred in, respectively, 100% and 20% of the intervals. Neither statement reflects that there were five occurrences of the behavior and that these five occurrences occupied 50% of the observation period.

This analysis demonstrates that the response dimensions of frequency and duration are confounded in interval sampling. Interval measurement errors in describing the frequency of behavior not only because of multiple responses within intervals (Altmann and Wagner, 1970; Mitchell, 1968), but also, because of responses that endure across successive intervals. This study showed that interval measurement errors in describing the duration of a behavior as a function of the time per response of the per response of a low duration, high frequency behavior will be sufficiently small that excessive measurement error will occur even if the observation interval is extremely brief.

The procedural deficiency that produces the error in interval sampling is easily identified. An entire interval is (1) scored no matter how brief the behavioral example within that interval, or (2) not scored if the behavior does not occur throughout the interval. This procedure guarantees that the sampling results will overestimate (1 above) or underestimate (2 above) the true extent of the observed behavior. Specification of a minimum duration of behavior that must be observed within an interval to score that interval does not solve the problem. Requiring that an example of the behavior occupy at least one-half of an interval to score the interval would be equivalent to conducting whole interval time sampling with the length of the observation interval halved. The net result will be to go from an overestimation to an underestimation. In summary, interval time sampling as a means of estimating behavioral duration seems to have little to recommend it. The resultant error is biased, and is likely to be both excessive and inconsistent.

For momentary time sampling, when observations were conducted each 5, 10, 20, and 60 sec, the per cent of observations scored agreed closely with the per cent of time the behavior occurred. This close agreement was evidenced across changes in the continuous measure. When observations were conducted each 120 sec, there were sessions where the difference between the momentary and continuous measure approached 20%. When observations were conducted each 200 sec, there was a session where this difference reached 40%. In our previous study (Powell, *et al.*, 1975), where the sessions were 20-min long, differences of 20% or more were first observed when the momentary time samples were conducted each 240 sec. These preliminary findings indicate that momentary time samples must be conducted surprisingly frequently for the sample results to mirror consistently the true state of nature, *i.e.*, a continuous measure of the behavior.

A comparison of the momentary and interval sampling results showed that momentary time sampling was superior in estimating the duration the behavior occurred. In addition to being a more accurate measurement procedure, momentary time sampling is more easily accomplished than interval time sampling. It does not require constant surveillance, but only a periodic assessment as to whether the behavior is, or is not, occurring. Momentary time sampling is, in short, the measurement procedure that should be employed in investigations where duration is the response dimension of interest.

REFERENCES

Altmann, S. A. and Wagner, S. S. Estimating rates of behavior from Hansen frequencies. *Primates,* 1970, **11**, 181-183.

Arrington, R. Time sampling in studies of social behaviours: a critical review of techniques and results with research suggestions. *Psychological Bulletin,* 1943, **40**, 81-124.

Barrish, H., Saunders, M., and Wolf, M. Good behavior game: effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis,* 1969, **2**, 119-124.

Hall, R. *Behavior modification: the measurement of behavior.* Lawrence, Kansas: H and H Enterprises, 1971.

Hauserman, N., Walen, S., and Behling, M. Reinforced racial integration in the first grade: a study in generalization. *Journal of Applied Behavior Analysis,* 1973, **6**, 193-200.

Hutt, S. and Hutt, C. *Direct observation and measurement of behavior,* Springfield, Illinois: Charles C Thomas, 1970.

Milby, J. Modification of extreme social isolation by contingent social reinforcement. *Journal of Applied Behavior Analysis,* 1970, **3**, 149-152.

Mitchell, A. D. Intercorrelations of maternal and infant behaviors in *Macaca mulatta. Primates,* 1968, **9**, 85-92.

Powell, J., Martindale, A., and Kulp, S. An evaluation of time-sample measures of behavior. *Journal of Applied Behavior Analysis,* 1975, **8**, 463-469.

Received 23 January 1976.
(Final acceptance 1 July 1976.)